Integration of Postgenomic Data for GMA to Simulate a Metabolic Circuit

Takayuki Tanaka c673054t@bse.kyutech.ac.jp
Hiroyuki Kurata kurata@bse.kyutech.ac.jp

Department of Biochemical Engineering and Science Kyusyu Institute of Technology, 680-4 Kawazu, Iizuka, Fukuoka 820-8502, Japan

Keywords: GMA formulation, real-coded genetic algorithm, multiple linear regression analysis, sensitivity analysis, clustering

1 Introduction

In order to understand cell metabolism as a system, it is important to formulize its molecular interaction network and to analyze it in the system level. In the postgenomic era, various experimental data such as transcriptome, proteome, and metabolome have accumulated rapidly. Such information is greatly available for constructing a molecular interaction map and simulating its dynamic features. We aim at developing the strategy that integrates such postgenomic data efficiently for simulating a metabolic system. General Mass Action (GMA) is a powerful candidate that integrates the postgenomic data [1], because GMA readily employs the profile of metabolite concentrations, fluxes, and enzyme concentrations. Then, a problem is that it is difficult to measure the kinetic parameters *in vivo* based on the above technologies. Generally, the measurement of a kinetic parameter requires the extensive and laborious experiments based on the kinetic theory. Therefore, we present a novel strategy that connects the postgenomic data to the GMA simulation.

2 Method

In the GMA formulation, the dynamics of each dependent variable depends on potentially all other dependent variables as well as on a set of independent variables. In order to build a dynamic model of metabolic networks, we employed GMA:

$$\frac{\mathrm{d}X_i}{\mathrm{d}t} = \sum_{k=1}^{P} \alpha_{ik} \prod_{j=1}^{n} X_j^{g_{ijk}} - \sum_{k=1}^{Q} \beta_{ik} \prod_{j=1}^{n} X_j^{h_{ijk}} \qquad (i = 1, \ldots, N), \tag{1}$$

where α_{ik} and β_{ik} are rate constants that characterize the flux rates between pools or variables, and g_{ijk} and h_{ijk} are the kinetic orders.

The problem is that most of the values for those kinetic parameters remain to be determined, because it is hard to measure all the parameters that vary *in vivo* with time and environmental changes. As a method to overcome such a problem, we present the method that combines the real-coded genetic algorithms (RCGAs) with the multiple linear regression analysis. The multiple linear regression analysis can be applied to each term of GMA that corresponds to the flux, solving the kinetic parameters, greatly reducing the number of the parameters to estimate. The use of RCGAs estimates the values of the remaining parameters. We used anaerobic fermentation pathway in *Saccharomyces cerevisiae* as a verification model. The CADLIVE (Computer-Aided Design of LIVing systems) Simulator was used as a numerical computation tool, and the statistic analysis software of SAS as clustering tool.

3 Results and Discussion

As search parameters of RCGAs, we chose 10 parameters, which consisted of the rate constants and the kinetic orders in GMA, and explored the 10-dimensional parameter space to fit the simulated result to the experimental data. Consequently, 20 sets of the parameter combinations were obtained as optimal ones. It turned out that the parameter values of each combination were distributed among them. Therefore, in order to further explore the parameter combinations, we performed sensitivity analysis regarding the sets of the parameters by using the CADLIVE system, as shown in Fig. 1.

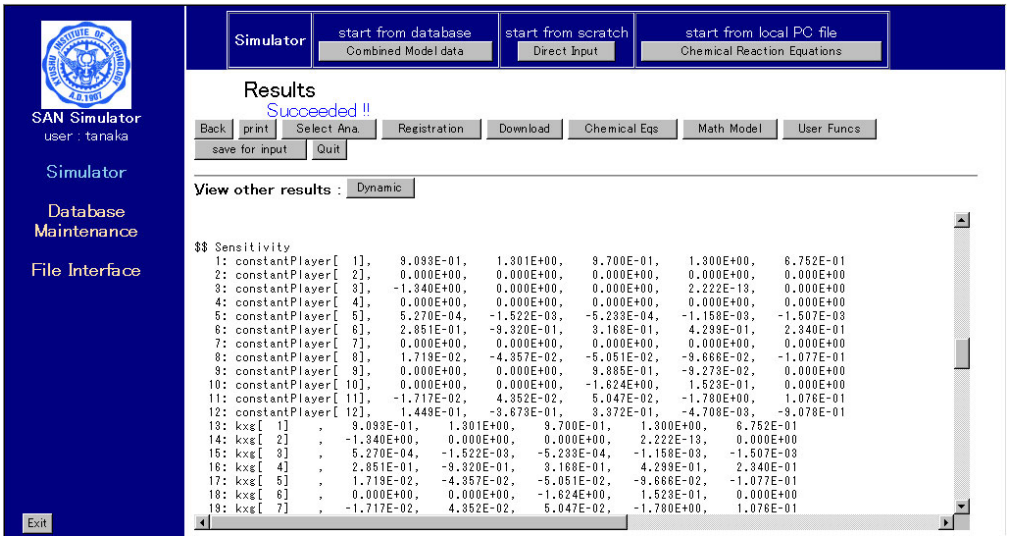

Figure 1: Dynamic simulator with CADLIVE (sensitivity analysis).

The class average method in the statistic analysis software of SAS was employed to classify the distribution pattern of the parameter combinations, as shown in Fig. 2. It turned out that it was roughly divided into two clusters, one contains the parameter set that shows the highest robustness, and the other does not contain it. The former showed a higher robust property than the latter. Since the important part of the metabolic circuit is made robustly, it may be quite reasonable to select a robust circuit. Therefore, it is important to grasp the properties of the system by analyzing the sensitivity.

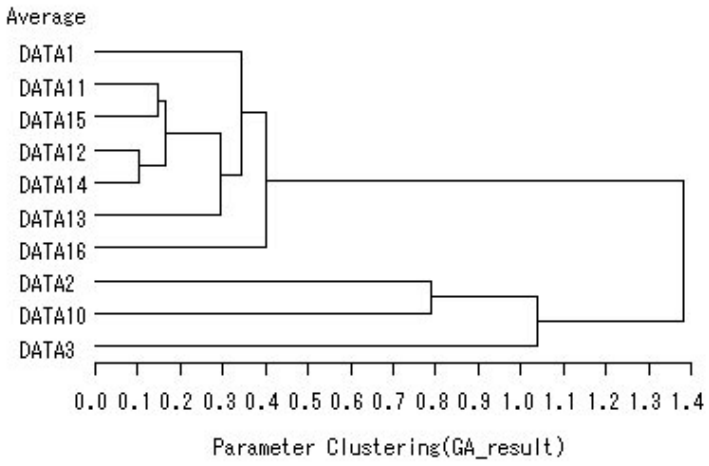

Figure 2: Clustering for the parameter sets.

References

[1] Kurata, H. and Taira, K., Two-phase partition method for simulating a biological system at an extremely high speed, *Genome Informatics*, 11:185–195, 2000.

[2] Kurata, H., Matoba, N., and Shimizu, N., CADLIVE for constructing a large-scale biochemical network based on a simulation-directed notation and its application to yeast cell cycle, *Nucleic Acid Res.*, 31(14):1–15, 2003.

[3] Voit, E.O., *Computational Analysis of Biochemical Systems*, Cambridge University Press, 2000.